Radiology

ORIGINAL RESEARCH ■ BREAST IMAGING

Does Reader Performance with Digital Breast Tomosynthesis Vary according to Experience with Two-dimensional Mammography?[1]

Lorraine Tucker, DCR
Fiona J. Gilbert, FRCR
Susan M. Astley, PhD
Amanda Dibden, MSc
Archana Seth, FRCR
Juliet Morel, FRCR
Sara Bundred, MD
Janet Litherland, FRCR
Herman Klassen, FRCR
Gerald Lip, FRCR
Hema Purushothaman, FRCR
Hilary M. Dobson, FRCR
Linda McClure, BSc
Philippa Skippage, FRCR
Katherine Stoner, FRCR
For the Group

Purpose: To assess whether individual reader performance with digital breast tomosynthesis (DBT) and two-dimensional (2D) mammography varies with number of years of experience or volume of 2D mammograms read.

Materials and Methods: After written informed consent was obtained, 8869 women (age range, 29–85 years; mean age, 56 years) were recruited into the TOMMY trial (A Comparison of Tomosynthesis with Digital Mammography in the UK National Health Service Breast Screening Program), an ethically approved, multicenter, multireader, retrospective reading study, between July 2011 and March 2013. Each case was read prospectively for clinical assessment and to establish ground truth. A retrospective reading data set of 7060 cases was created and randomly allocated for independent blinded review of *(a)* 2D mammograms, *(b)* DBT images and 2D mammograms, and *(c)* synthetic 2D mammograms and DBT images, without access to previous examinations. Readers (19 radiologists, three advanced practitioner radiographers, and two breast clinicians) who had 3–25 (median, 10) years of experience in the U.K. National Health Service Breast Screening Program and read 5000–13 000 (median, 8000) cases per annum were included in this study. Specificity was analyzed according to reader type and years and volume of experience, and then both specificity and sensitivity were analyzed by matched inference. The median duration of experience (10 years) was used as the cutoff point for comparison of reader performance.

Results: Specificity improved with the addition of DBT for all readers. This was significant for all staff groups (56% vs 68% and 49% vs 67% [$P < .0001$] for radiologists and advanced practitioner radiographers, respectively; 46% vs 55% [$P = .02$] for breast clinicians). Sensitivity was improved for 19 of 24 (79%) readers and was significantly higher for those with less than 10 years of experience (91% vs 86%; $P = .03$) and those with total mammographic experience of fewer than 80 000 cases (88% vs 86%; $P = .03$).

Conclusion: The addition of DBT to conventional 2D screening mammography improved specificity for all readers, but the gain in sensitivity was greater for readers with less than 10 years of experience.

[1] From the Department of Radiology, University of Cambridge, Cambridge Biomedical Campus, Cambridge CB2 0QQ, England (L.T., F.J.G.); Department of Imaging Science and Biomedical Engineering, University of Manchester, Manchester, England (S.M.A.); Centre for Cancer Prevention, Wolfson Institute of Preventive Medicine, Queen Mary University of London, London, England (A.D.); West of Scotland Breast Screening Service, Glasgow, Scotland (A.S., J.L., H.M.D., L.M.); Department of Radiology, King's College Hospital, London, England (J.M.); Department of Radiology, University Hospital of South Manchester, Manchester, England (S.B.); North East Scotland Breast Screening Centre, Aberdeen, Scotland (H.K., G.L.); Department of Radiology, Imperial College Healthcare NHS Trust, London, England (H.P.); and Jarvis Breast Centre, Guildford, England (P.S., K.S.). Received October 24, 2015; revision requested January 6, 2016; revision received August 24; accepted September 29; final version accepted December 8. **Address correspondence to** F.J.G. (e-mail: *fjg28@cam.ac.uk).* The complete list of authors and affiliations is at the end of this article.

Study supported by National Institute for Health Research Health Technology Assessment Programme (09/22/182).

The addition of digital breast tomosynthesis (DBT) to two-dimensional (2D) mammography improves specificity and sensitivity in breast cancer screening. The Oslo Tomosynthesis Screening Trial investigators (1) reported a 27% increase in overall cancer detection rate and an estimated 15% decrease in false-positive recall rate with the addition of DBT. The Screening with Tomosynthesis or standard Mammography trial (2) showed a similar improvement in cancer detection, with an estimated 17.2% decrease in recall rate. Friedewald and colleagues' recent retrospective time series analysis (3), comparing performance before and after the introduction of DBT to standard 2D screening mammography, reported a 29% increase in cancer detection rate and an 11% reduction in recall rates. However, the results were not consistent throughout the 13 U.S. sites that participated in the study; some sites showed an increase in recall rates.

It is not clear whether the addition of DBT to standard 2D mammography will improve the diagnostic performance of all readers. Evidence from reader performance studies in 2D mammography suggests that the number of years of experience and the number of screening mammograms read per year affect diagnostic accuracy. In an Australian study, Rawashdeh et al (4) reported better performance in reading 2D mammograms for readers with high volumes (≥5000 per annum) compared with readers with low volumes and fewer years of experience. A U.S. study (5) compared the diagnostic performance of readers with the volume of mammograms read and the number of years reading mammograms; the results showed that readers reading fewer than 5000 mammograms per annum had the highest median recall rate. In the United States, readers must read at least 960 mammograms every 2 years (6), whereas in the United Kingdom, readers in the National Health Service Breast Screening Program must read at least 5000 mammograms per annum and undergo annual performance, or PERFORMS, testing (7).

The issue of training of radiologists in breast imaging has been raised in two U.S. studies. Miglioretti et al (8) compared the sensitivity and specificity of screening mammography among fellowship-trained and nonfellowship-trained radiologists during the course of their careers. They found that both groups continued to improve throughout their careers, although those without fellowship training showed significant improvement in the first 3 years of clinical experience, whereas those with fellowship training achieved desirable clinical standards within 1 year. Following up on this, Elmore et al (9) looked at the characteristics associated with performance in screening mammography. They found wide variability in sensitivity (74.5%–92.3%) even among radiologists with similar false-positive rates. Fellowship training in breast imaging was the only significant characteristic associated with improved sensitivity.

A multireader study, conducted in the United States, comparing 2D mammography with DBT-2D mammography showed improved performance for readers with a range of experience, measured by recall rate and receiver operating characteristic analysis. Rafferty et al (10) reported that diagnostic accuracy increased and recall rates for noncancerous cases were significantly decreased with the addition of DBT. Sensitivity increased for 26 of 27 radiologists and was statistically significant for 10 of them. However, Wallis et al (11) in the United Kingdom reported that the addition of DBT improved performance only of readers with less than 10 years of experience with 2D screening mammography. A recent study (12) in which the authors looked at readers with different levels of experience with DBT and compared their performance with the use of 2D mammography and that with use of DBT-2D mammography found that readers of all levels of DBT experience

Advances in Knowledge

- The addition of digital breast tomosynthesis (DBT) to two-dimensional (2D) mammography shows greater improvement in sensitivity for readers with less than 10 years of experience (91% vs 86%; $P = .03$) or for those who have read fewer than 80 000 cases (88% vs 86%; $P = .03$).
- With the addition of DBT to 2D mammography, specificity was improved for all readers regardless of length of mammographic reading experience (66% vs 55% [$P < .0001$] for those with <80 000 reads; 67% vs 53% [$P < .0001$] for those with ≥80 000 reads).
- The improvement in specificity was significant for radiologists and advanced practitioner radiographers ($P < .0001$ for both groups) and breast clinicians ($P = .02$).
- Readers with less experience tended to find more asymmetries and distortions with the addition of DBT to 2D mammography than with 2D mammography alone (87% vs 78%), but the difference was not statistically significant ($P = .1$).

Implications for Patient Care

- The addition of DBT to 2D mammography will result in improved cancer detection, especially when cases are read by less experienced readers.
- The addition of DBT to 2D mammography will result in fewer unnecessary recalls when cases are read by readers of all levels of experience.

Published online before print
10.1148/radiol.2017151936 **Content code:** BR

Radiology 2017; 283:371–380

Abbreviations:
DBT = digital breast tomosynthesis
2D = two-dimensional

Author contributions:
Guarantor of integrity of entire study, F.J.G. The complete list of author contributions is at the end of this article.

Conflicts of interest are listed at the end of this article.

performed similarly with DBT-2D mammography. However, the authors of this study did not take into account the readers' experience in background 2D mammography. Another study of eight radiologists with 3–13 years of mammographic screening experience showed that most readers achieved an improved cancer detection rate and/or a decreased false-positive rate, but there was no correlation with experience (13).

There is a need to establish whether the addition of DBT is equally effective in improving performance of readers with different mammographic experience. In this study we assessed whether the improved reader performance seen overall with the addition of DBT to 2D mammography (14,15) varies with reader experience in reading 2D mammograms.

Materials and Methods

Support for this study was given by Hologic (Bedford, Mass), which created synthetic 2D images for the reading study. The authors had control of all data in the study and the information submitted for publication. Scotland A Research Ethics Committee approved the study, and women gave written informed consent to participate in the TOMMY trial (A Comparison of Tomosynthesis with Digital Mammography, in the U.K. National Health Service Breast Screening Programme [14]).

Results from the 7060 cases in the reading data set have been reported previously (15). That prior article was an investigation of the sensitivity and specificity of DBT for detecting breast cancer subgroups, whereas in our study, we analyzed individual reader performance according to reader type and years and volume of experience.

Study Design and Participants

The TOMMY trial was a retrospective reading study designed to compare the diagnostic performance of 2D mammography with that of DBT-2D mammography and of DBT synthetic 2D mammography. Data were collected between July 2011 and March 2013 in six U.K. National Health Service Breast Screening Program centers from 8869 women (age range, 29–85 years; mean age, 56 years) recalled to the assessment clinic for a mammographic abnormality found at routine 2D screening mammography. Screening cases were read with 2D mammography alone by two readers and were recalled for assessment if both readers found an abnormality. When readers disagreed, the cases were subjected to further discussion and the decision to recall was made on the basis of a consensus opinion. Women younger than 50 years with a family history of breast cancer and who underwent annual mammographic screening were also included in the main trial but were not included in this analysis to avoid further complexity due to case heterogeneity. For details of family history cases, see Gilbert et al (14). Patients with breast implants, those who were pregnant, and those unable to give written informed consent were excluded. Only one case per woman was included in the study (14).

Image Acquisition

Participants underwent standard two-view 2D mammography of both breasts and two-view DBT imaging as a single procedure at the same breast compression with a digital mammography unit (Selenia Dimensions; Hologic, Bedford, Mass).

Readers

TOMMY trial readers were radiologists (n = 25), advanced practitioner radiographers (n = 4), or breast clinicians (n = 2). Readers had 3–25 (median, 10) years of mammographic reading experience in the U.K. Breast Screening Program, reading between 5000 and 13000 (median, 8000) examinations each year (Table 1). Each reader attended an intensive 1-day DBT workshop and read a test set of 80 cases (14). All readers participated in the prospective data collection and gained experience in DBT by reviewing more than 100 DBT cases at their own center before they started the retrospective reading study.

Five readers participated only in prospective reading (four radiologists and one advanced practitioner), and two retrospective readers (both radiologists) were excluded because they had read fewer than 200 cases in each arm. This left 24 readers (19 radiologists, three advanced practitioner radiographers, and two breast clinicians) to be included in the analysis. Images were viewed on workstations (SecurView DW, Hologic) optimized to read both 2D mammograms and DBT images.

Retrospective Reading Study

Each case was read by three different readers from three different centers: one using 2D mammography alone, one using DBT-2D mammography, and another using DBT–synthetic 2D mammography. A randomization program managed by Cambridge Clinical Trials Unit assigned weekly reading sets, comprising a mix of normal, benign, and cancer cases. Cases were randomly selected from the study data set to be distributed among readers and centers to minimize bias. Each reading set comprised approximately 40 cases per reader per week. Readers were blinded to the outcome status of each case, did not review any cases that originated from their own center, and read cases independently without access to prior examinations. A decision to recall the case was made, and the location, size, and type of any suspicious abnormality were recorded along with a five-point suspicion score (1 = normal, 2 = benign; 3 = probably benign, 4 = probably malignant, and 5 = malignant).

Statistical Analysis

It was not possible to identify individual reader opinion for family history cases because of the manner in which the prospective data were collected, so we used data from the 6020 assessment cases (of which 1158 were cancers). Of these, for the 24 readers who read at least 200 cases in both the 2D mammography and DBT-2D mammography arms of the TOMMY trial, we had 2D mammography reading and experience data on 5888 cases (1141 cancers). We had the corresponding DBT-2D

Table 1

Summary of Reader Experience

Site/Reader Code	Reader Type*	Years Reading Mammograms	Mammograms per Year	Years Reading Digital Mammograms
Site 1				
1†	1	6	7000	2
2	1	17	6000	4
3	1	11	6000	5
4	1	4	6000	4
5†	1	15	6000	4
Site 2				
6‡	1	3	10 000	3
7†	1	15	8000	12
8	1	4	10 000	4
9	1	4	8000	4
10†	2	8	5147	8
Site 3				
11	1	18	12 000	2
12	2	10	10 000	1
13	1	25	5000	5
14	2	7	13 000	1
15	1	10	6000	3
Site 4				
16	1	3.5	6000	3
17	1	10	7000	2
18	1	20	8000	3
19	3	7	10 000	2
20	1	18	10 000	1
Site 5				
21	1	17	9000	7
22	1	5	7000	5
23	1	22	7000	2
24	1	24	10 000	7
25	3	6	7000	6
Site 6				
26‡	1	25	13 994	2
27†	1	8	6549	2
28	1	6	8430	2
29	1	23	8061	2
30	1	3	8525	2
31	2	10	8000	2

* Reader type: 1 = consultant radiologist, 2 = advanced practitioner radiographer, 3 = breast clinician.

† Readers who only contributed to prospective reading and did not participate in the retrospective reading study.

‡ Readers who were excluded from analysis for insufficient retrospective reads.

mammography data for 5857 cases (1130 cancers). We did not analyze data from synthetic 2D mammographic images to assess reader performance as part of this study. For the purposes of this study, a true-positive case was one for which cancer was proven histologically by means of surgery or core biopsy. A true-negative case was one for which no cancer was found after a 3-year follow-up period.

We calculated sensitivities and specificities separately for readers below and above the median for years of experience, volume of mammograms read per year, and total experience (the product of years of experience and mammograms read per year). Formal significance testing in subgroups of experience to compare 2D mammography with DBT-2D mammography was performed by using the method of Duffy et al (16) to combine matched and unmatched data. This enabled us, for any given experience subgroup, to use all of the data rather than only data from cases for which both the 2D mammography and DBT-2D mammography readers were in the same subgroup. Thus, for example, for analysis of sensitivity to cancer for readers having a total experience of fewer than 80 000 mammograms (the median total experience), we were able to use all 612 cancers for which the 2D mammography readers were in this subgroup and all 553 cases for which the DBT-2D mammography readers had this experience, rather than only the 264 cases for which both 2D mammography and DBT-2D mammography readers had read fewer than 80 000 mammograms in their career. We examined the radiologic features of lesions that were detected with DBT-2D mammography compared with 2D mammography alone and related this to the level of experience of the reader. In particular, we calculated sensitivity within the three lesser-experience subgroups by dominant radiologic feature.

For formal comparison within staff categories, we carried out the same formal analysis as for the reader experience subgroup, with one exception. We performed unmatched analysis for advanced practice radiographers and breast clinicians by using conventional χ^2 tests, because only four cancers were read by both advanced practice radiographers and four were read by both breast clinicians. Statistical analysis was performed by using software (Stata version 10.0; Stata Corp, College Station, Texas). We interpreted *P* values less than .05 as indicating a significant difference.

Results

The characteristics of the study population are shown in Table 2. Of the 5888 cases in the 2D mammography arm,

Table 2

Participant Characteristics

Characteristic	Assessment Cases: No. Randomized*	Assessment Cases: No. of Cancers*
Age range		
<40 y	3 (<1)	1 (<1)
40–49 y	340 (6)	27 (2)
50–59 y	3568 (59)	462 (40)
60–69 y	1714 (29)	519 (45)
≥70 y	364 (6)	141 (12)
Not known	31	9
Breast density		
0%–24%	1636 (27)	378 (33)
25%–49%	2556 (43)	439 (38)
50%–74%	1376 (23)	271 (24)
75%–100%	396 (7)	63 (5)
Not known	56	8
Cancer type		
Invasive ductal (with or without DCIS)	…	788 (68)
Invasive lobular (with or without DCIS)	…	109 (9)
Invasive other (with or without DCIS)	…	59 (5)
DCIS	…	203 (18)
Cancer size		
Invasive cancers		
1–5 mm	…	73 (8)
6–10 mm	…	243 (26)
11–20 mm	…	434 (46)
21–50 mm	…	183 (19)
>50 mm	…	10 (1)
Unknown	…	13
DCIS		
1–5 mm	…	30 (15)
6–10 mm	…	30 (15)
11–20 mm	…	47 (24)
21–50 mm	…	78 (39)
>50 mm	…	15 (7)
Unknown	…	3
Cancer grade		
Invasive cancers		
1	…	242 (26)
2	…	504 (54)
3	…	180 (20)
Unknown	…	30
DCIS		
Low	…	10 (8)
Intermediate	…	31 (22)
High	…	97 (70)
Unknown	…	65
Lymph node status (invasive cancers only)		
Normal	…	514 (58)
<4 nodes positive	…	292 (33)
≥4 nodes positive	…	77 (9)
Unknown	…	73
Dominant radiologic feature		
Circumscribed mass	1814 (30)	145 (13)
Spiculated mass	712 (12)	508 (44)
Microcalcification	1006 (17)	282 (24)
Distortion	514 (8)	109 (9)
Asymmetric density	1837 (31)	107 (9)
None	137 (2)	7 (1)

Note.—DCIS = ductal carcinoma in situ.

* Percentages have been reported for known data. Data in parentheses are percentages.

4720 (80%) were read by radiologists, 654 (11%) by advanced practice radiographers, and 514 (9%) by breast clinicians. The corresponding numbers in the DBT-2D mammography arm were 4608 (79%), 704 (13%), and 505 (8%), respectively. The median number of years of experience was 10 (interquartile range, 5–18) in both arms. Median number of mammograms read per year was 8000 (interquartile range, 6000–10000) in both groups. Median total lifetime mammograms read (the product of years of experience and mammograms per year) were 70000 (interquartile range, 40000–153000) in the 2D mammography arm and 80000 (interquartile range, 40000–153000) in the DBT-2D mammography arm.

Table 3 shows the cases and cancers read, plus estimates of sensitivity and specificity for each individual reader, for 2D mammography and DBT-2D mammography. Figure 1 shows the corresponding sensitivities and specificities plotted for each reader. The upward gradient of most lines illustrates how the addition of DBT improved both sensitivity and specificity. Specificity was improved for all readers and sensitivity was improved for 19 of 24 (79%) readers. For

Table 3

Assessment Cases Read, Cancers, Sensitivity, and Specificity for 2D Mammography and DBT-2D Mammography according to Individual Reader

	No. of Assessment Cases Read*		No. of Cases Marked for Recall*		Sensitivity (%)		Specificity (%)		Sensitivity Difference (%)†	Specificity Difference (%)†
Reader Code	2D	DBT-2D	2D	DBT-2D	2D	DBT-2D	2D	DBT-2D		
2	253 (56)	307 (75)	86 (39)	106 (53)	70	71	76	77	1 (−15, 17)	1 (−8, 10)
3	254 (61)	239 (42)	142 (57)	109 (40)	93	95	56	65	2 (−8, 12)	9 (−1, 19)
4	318 (73)	275 (51)	136 (53)	106 (43)	73	84	66	72	11 (−4, 26)	6 (−3, 14)
8	447 (74)	305 (71)	239 (66)	119 (65)	89	92	54	77	3 (−7, 13)	23 (15, 31)
9	235 (43)	223 (53)	111 (33)	90 (45)	77	85	59	74	8 (−8, 24)	14 (4, 24)
11	230 (37)	263 (44)	114 (30)	119 (41)	81	93	56	64	12 (−3, 27)	8 (−2, 18)
12	229 (48)	259 (53)	135 (43)	135 (49)	90	94	49	58	4 (−7, 15)	9 (−1, 20)
13	218 (42)	245 (51)	141 (38)	93 (43)	90	84	41	74	−6 (−20, 8)	33 (23, 43)
14	215 (51)	257 (46)	133 (43)	101 (41)	84	89	45	72	5 (−9, 19)	26 (16, 37)
15	243 (29)	240 (40)	142 (27)	112 (38)	93	95	46	63	2 (−10, 14)	17 (7, 27)
16	248 (51)	184 (38)	135 (46)	96 (35)	90	92	55	58	2 (−10, 14)	3 (−8, 14)
17	195 (29)	222 (24)	113 (28)	121 (24)	97	100	49	51	3 (−4, 10)	2 (−9, 13)
18	264 (54)	314 (57)	156 (51)	151 (52)	94	91	50	61	−3 (−13, 7)	11 (2, 21)
19	249 (42)	263 (54)	165 (41)	163 (54)	98	100	40	48	2 (−3, 7)	8 (−2, 18)
20	210 (37)	213 (52)	149 (36)	136 (49)	97	94	35	46	−3 (−12, 6)	11 (0, 22)
21	249 (36)	257 (45)	138 (32)	118 (41)	89	91	50	64	2 (−12, 16)	13 (4, 23)
22	284 (63)	300 (57)	157 (57)	122 (53)	90	93	55	72	3 (−7, 13)	17 (8, 26)
23	191 (37)	171 (31)	87 (33)	71 (28)	89	90	65	69	1 (−14, 16)	4 (−7, 16)
24	250 (50)	209 (35)	138 (43)	101 (33)	86	94	53	61	8 (−5, 21)	8 (−2, 19)
25	265 (56)	242 (42)	151 (51)	118 (41)	91	98	52	62	7 (−2, 16)	9 (−1, 19)
28	214 (51)	212 (37)	107 (47)	73 (34)	92	92	63	78	0 (−12, 12)	14 (4, 25)
29	209 (42)	223 (45)	89 (31)	65 (34)	74	76	65	83	2 (−17, 21)	17 (8, 27)
30	208 (40)	206 (44)	90 (33)	81 (37)	83	84	66	73	1 (−15, 17)	7 (−4, 17)
31	210 (39)	228 (43)	114 (34)	89 (37)	87	86	53	72	−1 (−16, 14)	19 (8, 29)

Note.—2D = 2D mammography.

* Data in parentheses are number of cancers.

† Data in parentheses are 95% confidence intervals.

three readers, sensitivity was reduced by more than 1% with the addition of DBT. All three were experienced radiologists with more than 15 years of experience and with cumulative experience of reading more than 100 000 mammograms.

Also shown in Table 3 are the differences in sensitivity and specificity for individual readers, with 95% confidence intervals. The positive estimates again illustrate the improvement of both measures with the addition of DBT. This was generally statistically significant only for specificity, probably because of the relatively small number of cancer cases read by any individual reader.

Table 4 shows estimates of sensitivity and specificity according to the three experience measures and staff

Figure 1

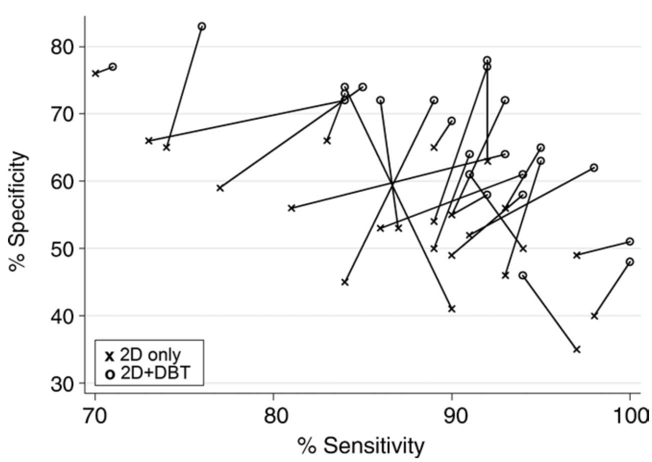

Figure 1: Graph shows combined change in sensitivity and specificity with the addition of DBT according to individual reader. An upward gradient from left to right indicates that both sensitivity and specificity are improved with DBT for that reader. Lines with a right-to-left upward gradient indicate increase in specificity but decrease in sensitivity for that reader. *2D* = 2D mammography

Table 4

Sensitivity and Specificity according to Reader Experience and Staff Category

Variable	No. of Readers	2D Mammography Sensitivity (%)	2D Mammography Specificity (%)	DBT-2D Mammography Sensitivity	DBT-2D Mammography Specificity
Years of experience					
<10	10	86 (470/544)	55 (1185/2139)	91 (448/493)	69 (1353/1974)
≥10	14	88 (534/609)	53 (1392/2615)	88 (571/647)	65 (1835/2817)
Mammograms per year					
<8000	10	86 (429/497)	56 (1111/1972)	88 (398/451)	67 (1318/1974)
≥8000	14	88 (575/656)	53 (1466/2782)	90 (621/689)	66 (1870/2817)
Total experience					
<80 000	12	88 (539/612)	55 (1399/2548)	92 (509/553)	66 (1557/2358)
≥80 000	12	86 (465/541)	53 (1178/2206)	87 (510/587)	67 (1631/2433)
Staff category					
Radiologist	19	86 (780/905)	56 (2125/3815)	88 (788/892)	68 (2515/3716)
Radiographer	3	87 (120/138)	49 (254/516)	89 (127/142)	67 (404/602)
Breast clinician	2	94 (92/98)	46 (192/416)	99 (95/96)	55 (223/409)

Note.—Data in parentheses are numerators and denominators.

Figure 2

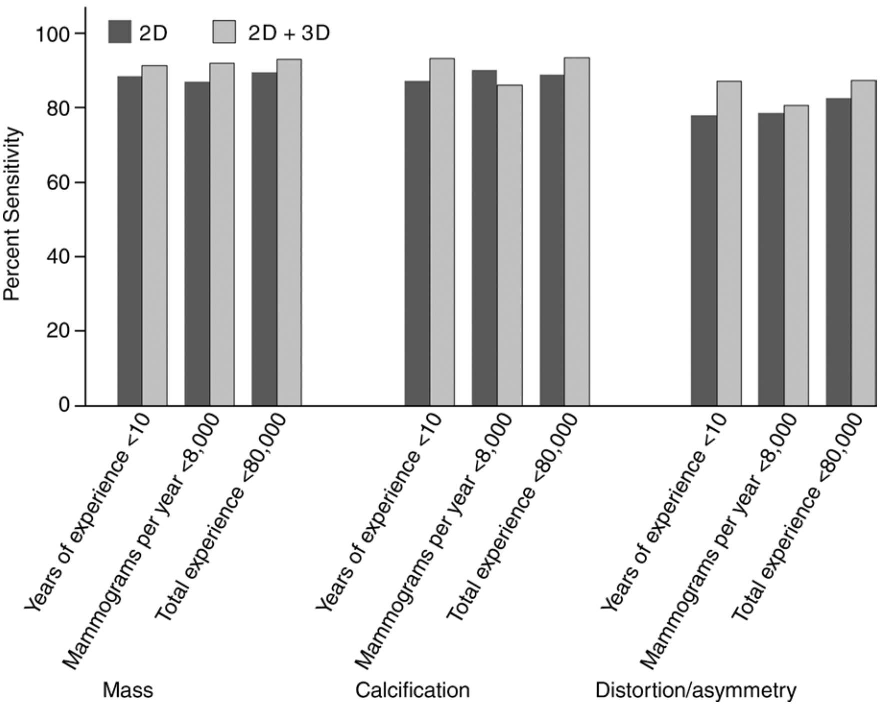

Figure 2: Bar graph shows sensitivity according to dominant radiologic feature for readers with less experience. *3D* = three dimensional.

category. For all subgroups, specificity for DBT-2D mammography was significantly higher ($P < .0001$ for radiologists and radiographers; $P = .02$ for breast clinicians). Sensitivity was significantly higher for DBT-2D mammography among those with less than 10 years of experience in reading mammograms ($P = .03$), for those with a total mammogram experience of fewer than 80 000 examinations ($P = .03$), and for radiologists ($P = .03$). There was suggestive evidence of higher sensitivity for breast clinicians ($P = .1$). No other significant or substantial nonsignificant differences were observed with respect to sensitivity.

As we noted previously, to better understand the sensitivity difference in those with less experience, we calculated sensitivity within the three lesser-experience subgroups according to dominant radiologic feature (Fig 2). Although no significant differences in sensitivity were observed in these small subgroups, the improvement conferred by DBT-2D mammography in readers with less than 10 years of experience was most striking where the dominant radiologic feature was asymmetric density or distortion (87% vs 78%).

Table 5 shows sensitivity according to imaging modality and years of reader experience for subgroups categorized by patient age, visually assessed breast density, histologic type of tumor, tumor size, and tumor grade. With 2D mammograms only, readers with less than 10 years of experience had lower sensitivity than did more experienced readers for patients younger than age 55 years, less than or equal to 50% density, invasive ductal carcinoma, invasive tumors of 10 mm or larger, and invasive grade

Table 5

Sensitivity according to Imaging Modality and Years of Experience in Subgroups of Age; Breast Density; and Tumor Type, Size, and Grade

	<10 Years of Experience		≥10 Years of Experience	
Variable	2D Mammography	DBT-2D Mammography	2D Mammography	DBT-2D Mammography
Age				
<55 y	83 (96/115)	93 (99/106)	91 (114/125)	88 (115/131)
≥55 y	87 (374/429)	90 (349/387)	87 (420/484)	88 (456/516)
Breast density (visual)				
<50%	87 (327/375)	90 (322/356)	88 (387/438)	88 (394/449)
≥50%	85 (143/169)	92 (126/137)	86 (147/171)	89 (177/198)
Tumor type				
Invasive ductal	86 (309/359)	92 (309/337)	88 (373/425)	89 (391/437)
Other	87 (161/185)	89 (139/156)	88 (161/184)	86 (180/210)
Invasive tumor size				
≤10 mm	84 (130/154)	87 (111/128)	87 (149/172)	85 (163/192)
>10 mm	88 (260/294)	93 (248/266)	88 (286/325)	92 (320/346)
Invasive tumor grade				
1–2	85 (309/363)	90 (276/305)	87 (327/374)	90 (380/421)
3	96 (72/75)	92 (72/78)	87 (95/109)	88 (92/104)

Note.—Data are percentages and numbers in parentheses are numerators and denominators.

1–2 tumors, but these differences were not statistically significant. For the less experienced readers, sensitivity was significantly improved with the addition of DBT for patients younger than 55 years (P = .048) and in tumors greater than or equal to 10 mm (P = .034). No other significant differences in sensitivity were observed between less and more experienced readers or between 2D mammography and DBT-2D mammography.

Discussion

The results of our study have shown a clear improvement in specificity for each individual reader. We also have demonstrated that the addition of DBT has a greater benefit in sensitivity for less experienced readers.

The improvement in specificity varied between 1% and 33% and was not related to volume of mammograms read each year or number of years of experience. In clinical practice, reductions in recall rate with the addition of DBT are also likely to vary according to background recall rate and recall/arbitration policy, as well as reader-to-reader variation.

We also have demonstrated a nonsignificant improvement in sensitivity for 80% of readers and a significant increase in sensitivity for readers with less than 10 years of experience with screening mammography. There was suggestive but nonsignificant evidence that the increased sensitivity of DBT-2D mammography for readers with lesser experience was largely due to distortions and asymmetric density as the main radiologic feature in patients. For example, for readers with less than 10 years of experience, when the main radiologic feature was distortion or asymmetry, sensitivity was 78% for 2D mammography versus 87% for DBT-2D mammography. These radiologic features are often subtle and are frequently missed, particularly by inexperienced mammogram readers. It is encouraging that DBT is especially useful for identification of these difficult features. There were no other major observations with respect to radiologic features. The lack of significant improvement in sensitivity overall (from 87% to 89%) in this study may be because the addition of DBT had minimal effect on the performance of this group of experienced, high-volume readers.

A strength of this study was that it was a very large reader study and included a large number of high-volume readers with a wide range of years of experience. The patients with cancer in this study were representative of those seen in the U.K. national Breast Screening Program, with similar radiologic feature distribution and type, grade, and size of cancer (17). A further strength was that cases were randomly allocated to the readers.

Weaknesses included the retrospective nature of the study. Therefore, readers knew that their decisions had no clinical implications, which may have resulted in them having less concern regarding missing a cancer. In addition, this was a highly enriched data set because most patients were those recalled to assessment as a result of a mammographic abnormality; therefore, the ratio of cancer cases to normal cases was higher than would be encountered in population screening and may not represent results obtained in a screening environment. A further weakness was that there were insufficient cancers to show an improvement in sensitivity in individual readers. Few cancers were detected with DBT alone because of the way the cases were collected; this introduced potential bias of increased specificity with DBT compared with that of 2D mammography. This could have resulted in overestimation of the true effect of the addition of DBT on specificity. The study design also meant that most patients with cancers were recalled after 2D screening mammography, and therefore, their cancers would have been visible at 2D mammography. This made it difficult to show increased sensitivity with the addition of DBT, particularly for experienced readers, and sensitivity with DBT was probably underestimated as a result of this method of data collection. In addition, because most patients were recruited from assessment clinics, this may have introduced a bias in the reader's decision to recall; that, in turn, would result in an increased recall rate, which would have affected sensitivity and specificity. Another limitation of the study was that readers did not have access to prior examinations, which they would have had in routine clinical practice. A randomized controlled study with recruitment at the point of screening is needed to ascertain the true potential clinical effects of our findings.

Relatively small numbers of cancer cases were read by breast clinicians because we had only two breast clinician readers in our study. Thus, the power to detect a significant difference in sensitivity for this category of staff was low, and the generalizability of this finding is limited. We could not consider reader clustering effects in the analyses because no readers read the same case with both imaging modalities and therefore matched sets (cases) were not nested for individual readers. As a result, true standard errors may have been larger than estimated and true significance levels may have been less extreme than estimated. However, the inaccuracy was likely to have been small within subgroups of readers according to experience.

We have demonstrated that the addition of DBT substantially and significantly improves specificity regardless of reader experience. The potential of DBT to reduce the number of false-positive recalls will be of considerable importance to screening programs, because this has been identified as a key factor deterring women from undergoing breast screening (18). Before DBT can be introduced into screening practice, the logistics of doing so must be considered. In particular, the effect of longer reading times on clinic workflow needs assessment. Reading times have been reported as being doubled with use of DBT, with a reading time of 91 seconds for DBT-2D mammograms compared with 45 seconds for 2D mammograms alone in a screening setting (1). However, this could be reasonable if recall rates are decreased and unnecessary diagnostic assessments are avoided (19). Cost implications and space required for data storage of DBT images are other areas that have been somewhat overlooked. Individual reading time varies considerably, and it is not known whether this is related to cancer detection.

Acknowledgments: The authors would like to acknowledge all who contributed to this study: Aberdeen: Karen A. Duncan (principal investigator, consultant radiologist), Jeanette Davidson (data manager), Maureen G.C. Gillan (research fellow); Cambridge: Richard Black (clinical scientist), Paula Willsher (research radiographer), Sridevi Nagarajan (senior data manager), Emily Dixon (data manager); Glasgow: Mabel Morrow (clinical specialist), Ann Mumby (superintendent radiographer), Julie Murphy (data manager); Guildford: Julie Cooke (principal investigator, consultant radiologist), Caroline Taylor (consultant radiologist), Victoria Cooke (associate specialist), Lindsay Mungutroy (data manager); King's College Hospital, London: Michael J. Michell (principal investigator, consultant radiologist), Asif Iqbal (data manager); Manchester: Mary Wilson (consultant radiologist), Jin Zhou (data manager), Elaine Harkness (postdoctoral research associate), Catriona Tate (data manager); St Bartholomew's Hospital, London: Ilyena Froud (consultant radiologist), Francis McInally (data manager); National Coordinating Centre for Physics of Mammography: Kenneth C. Young (professor), Celia Strudley (principal physicist).

Complete list of authors: Lorraine Tucker, DCR; Fiona J. Gilbert, FRCR; Susan M. Astley, PhD; Amanda Dibden, MSc; Archana Seth, FRCR; Juliet Morel, FRCR; Sara Bundred, MD; Janet Litherland, FRCR; Herman Klassen, FRCR; Gerald Lip, FRCR; Hema Purushothaman, FRCR; Hilary M. Dobson, FRCR; Linda McClure, BSc; Philippa Skippage, FRCR; Katherine Stoner, FRCR; Caroline Kissin, FRCR; Ursula Beetles, FRCR; Yit Yoong Lim, FRCR; Emma Hurley, FRCR; Jane Goligher, FRCR; Rumana Rahim, FRCR; Tanja J. Gagliardi, MD; Tamara Suaris, FRCR; Stephen W. Duffy, MSc.

Author affiliations: From the Department of Radiology, University of Cambridge, Cambridge Biomedical Campus, Cambridge CB2 0QQ, England (L.T., F.J.G.); Department of Imaging Science and Biomedical Engineering, University of Manchester, Manchester, England (S.M.A.); Centre for Cancer Prevention, Wolfson Institute of Preventive Medicine, Queen Mary University of London, London, England (A.D., S.W.D.); West of Scotland Breast Screening Service, Glasgow, Scotland (A.S., J.L., H.M.D., L.M.); Department of Radiology, King's College Hospital, London, England (J.M., J.G., R.R.); Department of Radiology, University Hospital of South Manchester, Manchester, England (S.B., U.B., Y.Y.L., E.H.); North East Scotland Breast Screening Centre, Aberdeen, Scotland (H.K., G.L.); Department of Radiology, Imperial College Healthcare NHS Trust, London, England (H.P.); Jarvis Breast Centre, Guildford, England (P.S., K.S., C.K.); Department of Clinical Radiology, Aberdeen Royal Infirmary, Aberdeen, Scotland (T.J.G.); and Department of Radiology, St Bartholomew's Hospital, London, England (T.S.).

Author contributions: Guarantor of integrity of entire study, F.J.G.; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; agrees to ensure any questions related to the work are appropriately resolved, all authors; literature research, L.T., F.J.G., S.M.A., H.P., H.M.D., J.G.; clinical studies, F.J.G., A.S., J.M., J.L., H.K., G.L., H.P., H.M.D., L.M., P.S., K.S., U.B., Y.Y.L., E.H., J.G., R.R., T.J.G., T.S.; statistical analysis, A.D., H.P., H.M.D., J.G., S.W.D.; and manuscript editing, L.T., F.J.G., S.M.A., S.B., H.K., G.L., H.P., H.M.D., L.M., J.G., T.J.G., T.S., S.W.D.

Disclosures of Conflicts of Interest: **L.T.** disclosed no relevant relationships. **F.J.G.** Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: grant/grant pending from Hologic. Other relationships: disclosed no relevant relationships. **S.M.A.** disclosed no relevant relationships. **A.D.** disclosed no relevant relationships. **A.S.** disclosed no relevant relationships. **J.M.** disclosed no relevant relationships. **S.B.** disclosed no relevant relationships. **J.L.** disclosed no relevant relationships. **H.K.** disclosed no relevant relationships. **G.L.** disclosed no relevant relationships. **H.P.** disclosed no relevant relationships. **H.M.D.** disclosed no relevant relationships. **L.M.** disclosed no relevant relationships. **P.S.** disclosed no relevant relationships. **K.S.** disclosed no relevant relationships. **C.K.** disclosed no relevant relationships. **U.B.** disclosed no relevant relationships. **Y.Y.L.** disclosed no relevant relationships. **E.H.** disclosed no relevant relationships. **J.G.** disclosed no relevant relationships. **R.R.** disclosed no relevant relationships. **T.J.G.** disclosed no relevant relationships. **T.S.** disclosed no relevant relationships. **S.W.D.** Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: grant from Philips. Other relationships: disclosed no relevant relationships.

References

1. Skaane P, Bandos AI, Gullien R, et al. Comparison of digital mammography alone and digital mammography plus tomosynthesis in a population-based screening program. Radiology 2013;267(1):47–56.

2. Ciatto S, Houssami N, Bernardi D, et al. Integration of 3D digital mammography with tomosynthesis for population breast-cancer screening (STORM): a prospective comparison study. Lancet Oncol 2013;14(7):583–589.

3. Friedewald SM, Rafferty EA, Rose SL, et al. Breast cancer screening using tomosynthesis in combination with digital mammography. JAMA 2014;311(24):2499–2507.

4. Rawashdeh MA, Lee WB, Bourne RM, et al. Markers of good performance in mammography depend on number of annual readings. Radiology 2013;269(1):61–67.

5. Cornford E, Reed J, Murphy A, Bennett R, Evans A. Optimal screening mammog-

raphy reading volumes; evidence from real life in the East Midlands region of the NHS Breast Screening Programme. Clin Radiol 2011;66(2):103–107.

6. American College of Radiology. Mammography Accreditation Program Requirements. http://www.acraccreditation.org/~/media/ACRAccreditation/Documents/Mammography/Requirements.pdf. Accessed May 8, 2015.

7. Wilson R, Liston J, eds. Quality assurance guidelines for breast cancer screening radiology. NHSBSP Publication No. 59. Sheffield, England: NHS Cancer Screening Programmes, 2011.

8. Miglioretti DL, Gard CC, Carney PA, et al. When radiologists perform best: the learning curve in screening mammogram interpretation. Radiology 2009;253(3):632–640.

9. Elmore JG, Jackson SL, Abraham L, et al. Variability in interpretive performance at screening mammography and radiologists' characteristics associated with accuracy. Radiology 2009;253(3):641–651.

10. Rafferty EA, Park JM, Philpotts LE, et al. Assessing radiologist performance using combined digital mammography and breast tomosynthesis compared with digital mammography alone: results of a multicenter, multireader trial. Radiology 2013;266(1):104–113.

11. Wallis MG, Moa E, Zanca F, Leifland K, Danielsson M. Two-view and single-view tomosynthesis versus full-field digital mammography: high-resolution X-ray imaging observer study. Radiology 2012;262(3):788–796.

12. Alakhras MM, Brennan PC, Rickard M, Bourne R, Mello-Thoms C. Effect of radiologists' experience on breast cancer detection and localization using digital breast tomosynthesis. Eur Radiol 2015;25(2):402–409.

13. Bernardi D, Caumo F, Macaskill P, et al. Effect of integrating 3D-mammography (digital breast tomosynthesis) with 2D-mammography on radiologists' true-positive and false-positive detection in a population breast screening trial. Eur J Cancer 2014;50(7):1232–1238.

14. Gilbert FJ, Tucker L, Gillan MG, et al. The TOMMY trial: a comparison of TOMosynthesis with digital MammographY in the UK NHS Breast Screening Programme—a multicentre retrospective reading study comparing the diagnostic performance of digital breast tomosynthesis and digital mammography with digital mammography alone. Health Technol Assess 2015;19(4):i–xxv, 1–136.

15. Gilbert FJ, Tucker L, Gillan MG, et al. Accuracy of digital breast tomosynthesis for depicting breast cancer subgroups in a UK retrospective reading study (TOMMY trial). Radiology 2015;277(3):697–706.

16. Duffy SW, Rohan TE, Altman DG. A method for combining matched and unmatched binary data. Application to randomized, controlled trials of photocoagulation in the treatment of diabetic retinopathy. Am J Epidemiol 1989;130(2):371–378.

17. National Health Service Cancer Screening Programmes. NHS Breast Screening Programme 2012 Annual Review. http://www.cancerscreening.nhs.uk/breastscreen/publications/2012review.html. Accessed August 19, 2015.

18. Kopans DB. Digital breast tomosynthesis from concept to clinical care. AJR Am J Roentgenol 2014;202(2):299–308.

19. Lee CI, Lehman CD. Digital breast tomosynthesis and the challenges of implementing an emerging breast cancer screening technology into clinical practice. J Am Coll Radiol 2013;10(12):913–917.